Journal of Experimental Psychology: Human Perception and Performance

Implicit Versus Explicit Frequency Comparisons: Two Mechanisms of Auditory Change Detection

Laurent Demany, Catherine Semal, and Daniel Pressnitzer

Online First Publication, September 20, 2010. doi: 10.1037/a0020368

Journal of Experimental Psychology:
Human Perception and Performance

0096-1523/10/$12.00 DOI: 10.1037/a0020368

Implicit Versus Explicit Frequency Comparisons: Two Mechanisms of Auditory Change Detection

Laurent Demany and Catherine Semal
Centre National de la Recherche Scientifique,
and Université de Bordeaux

Daniel Pressnitzer
Centre National de la Recherche Scientifique, Université Paris-Descartes, and Ecole normale supérieure, Paris, France

Listeners had to compare, with respect to pitch (frequency), a pure tone (*T*) to a combination of pure tones presented subsequently (*C*). The elements of *C* were either synchronous, and therefore difficult to hear out individually, or asynchronous and therefore easier to hear out individually. In the "present/absent" condition, listeners had to judge if *T* reappeared in *C* or not. In the "up/down" condition, the task was to judge if the element of *C* most similar to *T* was higher or lower than *T*. When the elements of *C* were synchronous, the up/down task was found to be easier than the present/absent task; the converse result was obtained when the elements of *C* were asynchronous. This provides evidence for a duality of auditory comparisons between tone frequencies: (1) implicit comparisons made by automatic and direction-sensitive "frequency-shift detectors"; (2) explicit comparisons more sensitive to the magnitude of a frequency change than to its direction. Another experiment suggests that although the frequency-shift detectors cannot compare effectively two tones separated by an interfering tone, they are largely insensitive to interfering noise bursts.

Keywords: auditory memory, pitch discrimination, sequence perception, temporal binding

Human listeners, as well as some animal species (Wright, Rivera, Hulse, Shyan, & Neiworth, 2000; Gentner, 2008), are sensitive to *relations* between successive sound stimuli, even in the presence of substantial inter-stimulus intervals (ISIs). This sensitivity is likely to depend at least in part on automatic, bottom-up neural mechanisms. In recent psychophysical studies (Demany & Ramos, 2005; Demany, Pressnitzer, & Semal, 2009; Demany, Semal, Cazalets, & Pressnitzer, 2010), we have obtained results supporting that view. The studies in question have suggested that, counterintuitively, two successive pure tones differing in frequency can evoke a percept of directional pitch change even when the first tone has not been consciously perceived. This has been taken as evidence for the existence of automatic "frequency-shift detectors" (FSDs) in the auditory system. Such detectors might be instrumental in auditory scene analysis (Bregman, 1990) by establishing perceptual links between temporally separate sounds emanating from the same acoustic source. In the present paper, we report data providing further support for the existence of FSDs. Moreover, we show that the "implicit" frequency comparisons made by the FSDs are fundamentally different from the "explicit" frequency comparisons that can be made by a human listener between two consciously audible tones.

In our three previous studies, listeners were presented on each trial with a sequence of two sounds: a "chord" of synchronous pure tones with randomly chosen frequencies, followed (or sometimes preceded) by a single pure tone (*T*). The tonal elements of the chord (varying in number from 5 to 10 across studies) were always sufficiently spaced in frequency to be resolved in the cochlea, and *T* was generally separated from the chord by an ISI of several hundreds of ms. There were two experimental conditions, called "present/absent" and "up/down." In the present/absent condition, *T* could be either identical to a randomly chosen element of the chord, or positioned halfway in log-frequency between two adjacent and randomly chosen elements; the listener's task was to indicate if *T* was present in the chord or absent from it. In the up/down condition, on the other hand, *T* was always slightly higher or lower (by one semitone, typically) than a randomly chosen element of the chord, and the listener's task was to indicate the direction of the corresponding frequency shift. In the present/absent condition, performance was generally very poor, thus suggesting that the elements of the chords could not be heard out individually. This was consistent with previous research: because the elements of the chords were gated on and off synchronously, they were grouped together, resulting in "informational masking" (Neff & Green, 1987; Neff, 1991, 1995; Durlach et al., 2003; Kidd, Mason, Richards, Gallun, & Durlach, 2008). Nevertheless, performance was quite good in the up/down condition. Listeners reported that in the latter condition, they could hear *T* as the end point of an ascending or descending pitch movement, subjectively starting from the chord as a whole rather than from one of its elements. These results can be accounted for by a simple FSD model assuming that: (1) the auditory system contains two subsets of automatic FSDs, respectively tuned to upward and downward frequency

Laurent Demany and Catherine Semal, Centre National de la Recherche Scientifique, and Université de Bordeaux; and Daniel Pressnitzer, Centre National de la Recherche Scientifique, Université Paris-Descartes, and Ecole Normale Supérieure, Paris, France.

Part of the results reported in this article were presented at the 14th International Symposium on Hearing (Cloppenburg, Germany, August 2006) and at Acoustics '08 (Paris, France, July 2008).

Correspondence concerning this article should be addressed to Laurent Demany, BP 63, Université de Bordeaux, 146 rue Leo Saignat, F-33076 Bordeaux, France. E-mail: laurent.demany@u-bordeaux2.fr

shifts; (2) within each subset, the FSDs are maximally sensitive to small shifts (the optimal shift magnitude being about 1/10 octave according to Demany et al., 2009); (3) Listeners may not have conscious access to the activation of a given FSD subset, but are sensitive to the difference between the strengths of the activations of the two subsets. Assumptions 1 and 2 are sufficient to account for the good performance obtained in the up/down condition. Assumption 3 implies that the FSDs will be unhelpful in the present/absent condition because on both "present" and "absent" trials, the two subsets of FSDs are expected to be activated with approximately the same relative strength.

Since the FSDs are supposedly able to detect a change in a tone that has not been consciously perceived, they represent an implicit, automatic change detection mechanism. A change detection mechanism of that kind is especially useful for hearing, more than for vision, as auditory information is by its very nature transient (Demany et al., 2010). In many circumstances, however, auditory change detection can be explicit. This happens, for instance, when listening to a musical melody, or a speech stream. The main aim of this paper is to compare implicit and explicit modes of auditory change detection.

In the first experiment reported below, we manipulated the audibility of the individual elements of pure-tone combinations. The present/absent and up/down tasks were performed using both chords of synchronous pure tones and, in the place of chords, sets of asynchronous pure tones. Due to their asynchrony, the elements of the latter tonal sets could be individually heard out more easily than the elements of the chords (Neff, 1995; Durlach et al., 2003). It could thus be expected that, with these new stimuli, the present/absent task would become easier than the up/down task. We reasoned that such a trend reversal would provide compelling evidence for the existence of two basically different types of frequency comparisons between successive tones: on the one hand implicit comparisons made by automatic FSDs, and on the other hand explicit comparisons between conscious pitch percepts.

Experiment 1 also served to test a possible interpretation of our previous results that did not call upon automatic FSDs. It could be hypothesized that listeners' poor performance in the present/absent condition of our previous studies was not due to an inability to hear out the elements of the chords, but was caused instead by an inability to use, in the present/absent task, an appropriate and stable decision criterion (see in this respect Demany & Ramos, 2005). From the point of view of signal detection theory (Green & Swets, 1974), the judgments required in the present/absent task are more demanding in terms of criterion setting than the judgments required in the up/down task. In order to test the idea that the poor performance observed in the present/absent condition stemmed from criterion-setting difficulties rather than from an inability to hear out the chords' elements, Demany and Ramos (2005) ran a "present/close" condition in which, on each trial, the *T* tone was either identical to one element of the chord (as in the present/absent condition) or positioned at random slightly (1.5 semitone) above or below one element. In terms of criterion setting, the present/close task was even more demanding than the present/absent task since in the former task the frequency shifts to be detected were smaller. Nevertheless, the easier of these two tasks appeared to be the present/close task. This outcome was consistent with the FSD model described above, and supported the idea that the elements of the chords were not individually audible. Unfortunately, few data were collected in the present/close condition that we just described. It remained to be checked that the present/absent task was more difficult than the up/down task because the elements of the chords could not be heard out individually rather than because of an intrinsic difficulty of present/absent judgments. This goal is achieved by the first experiment reported here.

Our second experiment was concerned with the nature of the comparisons made by the FSDs. In our previous studies, the *T* tone was usually presented after the chord rather than before it. When this was the case, one possible explanation for listeners' success in the up/down condition was that the representation of the *T* tone in the auditory system was not the same on "up" and "down" trials, due to a differential forward masking effect of the chord element closest to *T* in frequency. In other words, the idea is that thanks to some property of forward masking, the correct response to be given on an "up" or "down" trial could be guessed on the basis of the percept evoked by *T* alone, without genuinely *comparing* this tone to the most similar element of the chord. This hypothesis seems implausible because a sound has no effect on the detection threshold of a subsequent sound when the ISI exceeds about 200 ms (Zwislocki, Pirodda, & Rubin, 1959; Elliott, 1971), whereas the up/down task remains easier than the present/absent task for much longer ISIs (Demany & Ramos, 2005). However, even though in our previous studies the chord was generally unable to raise the detection threshold of *T*, the possibility of subtle suprathreshold forward masking effects could not be ruled out a priori. Such effects, if present, were also liable to play an important role when *T* was presented before the chord rather than after it: In that case, *T* might have affected the timbre of the chord. We reasoned that any perceptual cues introduced by forward masking in the up/down task should be lost if, on both "up" and "down" trials, a loud burst of wideband noise were presented between the chord and *T*: Since the final sound in the sequence should be affected much more by the noise than by the sound preceding the noise (if any forward masking effect occurs), the final sound should be affected in the same way on "up" and "down" trials.

Experiment 1

Method

On each trial in this experiment, the listener had to compare a pure tone (*T*) to a combination of five pure tones presented subsequently (*C*). Each tone had a total duration of 300 ms, was gated on and off with 20-ms raised-cosine amplitude ramps, and had a nominal sound pressure level (SPL) of 65 dB. The five elements of *C* were spaced in frequency by equal intervals of 550 musical cents (1 cent = 1/100 semitone = 1/1200 octave), and *C* was randomly positioned between 125 and 4000 Hz, using a logarithmic frequency scaling. As in our previous studies described above, the listener had to perform either a present/absent task or an up/down task. In the present/absent condition, *T* was either identical to a randomly selected element of *C* (the five possible choices being equiprobable) or positioned halfway in log-frequency between two adjacent elements (four possible choices); the correct response was "present" in the former case, and "absent" in the latter case. In the up/down condition, *T* was positioned at random 100 cents above or below a randomly selected element of *C*; the listener had to identify the direction of the frequency shift taking place from *T* to the most similar element of *C*.

C was of three possible types, depending on the temporal relations of its elements. In one case, the elements were synchronous, so that *C* was a chord. In a second case, the five elements were presented successively rather than simultaneously, with stimulus-onset-asynchronies (SOAs) of 100 ms between consecutive elements and a random ordering of the five frequencies. The third case was identical to the second, except that this time the SOAs had a duration of 250 ms. When *C* was a chord (SOA = 0 ms), it was separated from *T* by a 1-s ISI. In the other two cases, the third element of *C*—that is, its median element in the time domain—was also separated from *T* by a 1-s ISI. As in our previous studies, a random melody of pure tones was presented at the beginning of every trial. This random melody, serving both as a warning signal and as a pitch eraser (see Demany & Ramos, 2005, footnote 1), consisted of five 300-ms tones with frequencies drawn from between 125 and 4000 Hz. These five tones were concatenated without any ISI, and were followed by *T* after a 600-ms silent ISI.

Trials were organized in blocks of 50, during which the SOA of the elements of *C* was fixed, as well as the listener's task (present/absent or up/down). Within a given experimental session, generally consisting of eight blocks of trials, the two tasks were performed alternately from block to block but the SOA did not change. A given SOA was used in two consecutive sessions; the ordering of the SOAs varied randomly across listeners. Overall, in the experiment proper, 400 trials per listener were run for each combination of SOA and task.

The stimuli were produced at a sampling rate of 44.1 kHz, using a 24-bit sound card (RME). They were presented binaurally (diotically), using Sennheiser HD265 headphones. The listener, seated in a triple-walled soundproof booth (Gisol, Bordeaux), gave his or her responses by means of mouse clicks on two virtual buttons. Response time was unlimited. There was no immediate feedback about response accuracy, but listeners were allowed to examine their results following each block of trials.

Five listeners with normal hearing (four men, one woman) were tested. This group included four students in their twenties and the first author (54 years). All of these listeners were amateur musicians. Only two of them had previous experience with psychoacoustics. For each listener, except the first author, the experiment proper was preceded by three or four training sessions. In the initial training session, listeners were first familiarized with the two tasks using chords consisting of only three pure tones, very widely spaced in frequency.

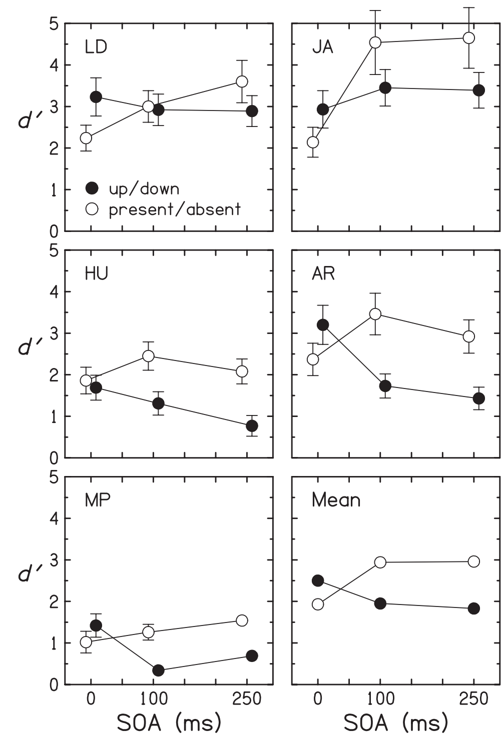

Figure 1. Individual and mean performance in Experiment 1, as a function of the perceptual task (up/down versus present/absent) and the stimulus onset asynchrony (SOA) of the elements of *C*. The SOA was equal to 0, 100, or 250 ms. Five listeners were tested (LD, JA, HU, AR, MP). For clarity, in the five corresponding panels, the data points have been slightly shifted horizontally, to the left for the present/absent task and to the right for the up/down task. The error bars represent 95 % confidence intervals around d'; these confidence intervals were computed as suggested by Macmillan and Creelman (1991).

Results

Performance was measured in terms of the sensitivity index d' of signal detection theory. The individual and mean data are displayed in Figure 1. When the elements of *C* were synchronous (SOA = 0 ms), listeners were usually more successful in the up/down task that in the present/absent task. Inversely, when the elements of *C* were asynchronous, listeners were usually more successful in the present/absent task that in the up/down task; however, increasing the SOA from 100 ms to 250 ms did not enhance appreciably the latter trend. Overall, making the elements of *C* asynchronous rather than synchronous improved performance in the present/absent task, but produced a trend in the opposite direction for the up/down task.

A repeated-measures ANOVA of the data displayed in Figure 1 (using Geisser-Greenhouse adjustments for the probability levels) confirmed the existence of a significant interaction between SOA and task, $F(2, 8) = 33.2$, $p = .0014$, $\eta^2 = 0.89$. A main effect of task was also found, $F(1, 4) = 9.7$, $p = .04$, $\eta^2 = 0.71$, but there was no main effect of SOA, $F(2, 8) < 1$. Tukey post-hoc tests indicate that d' was significantly larger in the up/down task than in the present/absent task when there was no SOA ($p = .026$), whereas the opposite was true for the 100-ms SOA ($p = .001$) and the 250-ms SOA ($p < .001$). Tukey tests also show that, in the present/absent task, d' was significantly smaller when there was no SOA than when the SOA was 100 ms ($p = .025$) or 250 ms ($p = .023$). In the up/down task, on the other hand, similar tests failed to reveal significant SOA effects.

Discussion

These findings clearly suggest that frequency comparisons between temporally separated tones can be made in two basically different ways. However, in order to interpret the data precisely, it was desirable to compare them to those expected from an ideal listener making explicit comparisons between C and T. We did so using minimal and standard assumptions with respect to the comparison mechanism itself. The aim was not to fit the data, but rather to circumscribe the possible relations between performance in the two tasks when it is assumed that the components of C can be heard out individually.

First, we tested the ideal listener model used by Demany and Ramos (2005). This ideal listener (Λ) is able to hear out individually all the elements of C, even when they are synchronous, but its performance is limited by a Gaussian random noise affecting the sensory encoding of the elements, independently of each other; T is supposed to be encoded with perfect accuracy. The random variable corresponding to the noise has a mean of 0 and a standard deviation of σ. In each task, Λ measures the distance between T and each element of C (corrupted by the noise) on a log-frequency scale, and then identifies the shortest of the five distances. In the up/down task, Λ responds "up" if the element of C that is identified as the closest to T is higher than T, and "down" otherwise. In the present/absent task, Λ responds "present" if the shortest of the five distances is smaller than 137.5 cents—that is, one fourth of the true distance between the elements of C—and the response is "absent" otherwise. (The criterion distance of 137.5 cents is somewhat arbitrary but we found that changing this parameter had little effect on d'.) We assessed the performance of Λ in the two tasks for five values of σ, ranging from 70 cents to 110 cents by steps of 10 cents. The results are plotted as squares in Figure 2. It can be seen that Λ is equally effective in the two tasks when σ is such that d' is about 2. For smaller values of σ, yielding better overall performance, Λ is more effective in the present/absent task than in the up/down task; this trend is reversed when σ is such that $d' < 2$ in both tasks.

In the model considered up to now, performance is limited by a noisy encoding of the elements of C, but not T. This was an appropriate model for the main experiment of Demany and Ramos (2005), in which the elements of C were always synchronous—a circumstance propitious to informational masking—and T was presented after C. In the current experiment, however, the elements of C were not always synchronous and T was presented before C. Since the memory trace of a tone can be strongly affected by the presentation of subsequent tones (Deutsch, 1972, 1999), it is reasonable to assume that T had a more noisy internal representation than the elements of C when these elements were asynchronous and successively compared to T. This led us to consider an ideal listener Λ' behaving exactly like Λ but hampered by noise in the encoding of T alone, not the elements of C. For the five values of σ previously selected, the results of Λ' in the two tasks are plotted as diamonds in Figure 2. It can be seen that the relation between performance in the two tasks is rather similar for Λ and Λ'.

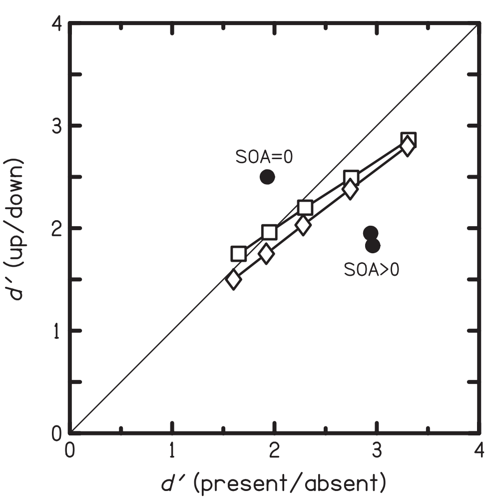

Figure 2. Theoretical versus observed relations between performance (d') in the present/absent and up/down tasks of Experiment 1. Open squares and diamonds represent the predictions of two models described in the text. Black disks represent the mean experimental results obtained when the elements of C were synchronous (SOA = 0) and when they were asynchronous (SOA > 0). In the latter case, the SOA was either 100 ms or 250 ms.

Neither of the models described above can be entirely correct since each of them assumes that some portion of the sound sequences presented to the listeners was encoded without any internal noise. In fact, both C and T were certainly corrupted by internal noise. However, each model is relevant because it represents a limiting case. We verified that any intermediate model makes predictions that lie in between those shown in Figure 2. Remarkably, our modeling shows that, with some magnitude of internal noise, one could expect to find no advantage of the present/absent task over the up/down task, even under the assumption that the elements of C are explicitly audible as separate entities.

Let us now compare the model predictions with the experimental data. The three black disks displayed in Figure 2 represent the mean results of Experiment 1, also plotted in the bottom-right panel of Figure 1. Consider first the results obtained when the SOA was 0 ms. In that case, d' was significantly larger in the up/down task than in the present/absent task. The obtained pair of d' values is clearly inconsistent with the predictions of the two models defined above, or any intermediate model. We take this discrepancy as evidence that, when the elements of C were synchronous, performance in the up/down task was not entirely determined by *explicit* comparisons between T and the elements of C. Instead, listeners presumably took advantage of *implicit* comparisons made by automatic FSDs, which were helpful in the up/down condition but less helpful or completely useless in the present/absent condition, as predicted by the FSD model that we have described in the Introduction.

In the present/absent condition, nevertheless, performance was good; it was not *much* poorer than in the up/down condition, in contrast to the findings of Demany and Ramos (2005, Experiment 1) or Demany et al. (2009, Experiment 1). This is probably due to the fact that, in the current experiment, T was presented before C, whereas the opposite was true in the experiments that we just mentioned. The presentation of T before C drew listeners' attention to the relevant spectral region of C, and thus gave listeners a certain amount of latitude for explicit pitch comparisons between T and an element of C, even when the elements of C were synchronous. Helmholtz (1859) had already pointed out that although the harmonics of a complex tone such as a piano note are normally not audible individually, it is possible to make a harmonic audible by presenting, just before the complex tone, another tone matched in frequency to the target harmonic. Interestingly, in most of the subsequent studies on human listeners' ability to hear out the spectral components of complex tones, the participants had to perform an up/down task rather than a present/absent task (Roberts & Bregman, 1991; Moore & Ohgushi, 1993; Bernstein & Oxenham, 2003, 2006, 2008; Moore, Glasberg, & Jepsen, 2009). This methodological choice may have been motivated in part by the fact that the up/down task was the easier one, as shown by the current experiment. However, our data lead us to question the validity of that choice: We argue that success in the up/down task did not necessarily imply that the target component of the complex tone could be explicitly heard out.

A preceding tone may also alter the timbre quality of a subsequent tone complex if it is matched to one component of the complex, potentially providing an additional cue to listeners (Darwin, 1984). However, this cue is likely to have been weak in the current experiment as it should have allowed excellent performance for "present" trials. This was not observed. The long ISI (1 s) used in the current experiment could explain the weakness of this timbre cue.

Consider now the experimental results that we obtained with nonzero SOAs. In that case, due to the SOAs themselves, each element of *C* was explicitly audible for some period of time. This facilitated, of course, explicit frequency comparisons between *T* and the elements of *C*. That is presumably the reason why performance in the present/absent task was better with nonzero SOAs than without such SOAs. On the other hand, the same SOA manipulation did not improve performance in the up/down task. One can account for the latter finding by assuming that implicit frequency comparisons based on FSDs are effective for immediately consecutive tones, but are less effective for tones separated by other tones. If this is true, the advantage provided by the FSDs in the up/down task was stronger when the elements of *C* were synchronous than when they were not synchronous. The corresponding deleterious effect of SOAs may have largely cancelled the advantage of SOAs for explicit comparisons.

We thus suggest that when the elements of *C* were not synchronous, the two tasks were performed mainly or exclusively by means of explicit comparisons between tones. Under that assumption, however, one would expect results consistent with those of the ideal listener Λ or Λ'. In fact, as shown by Figure 2, the model predictions were again definitely falsified; but this time, contrary to the trend observed when the elements of *C* were synchronous, the listeners tested in the experiment were unexpectedly successful in the *present/absent* task, given their performance in the up/down task. This trend is not so surprising since it is consistent with previous observations by Wickelgren (1969). In Wickelgren's study, listeners were presented with sequences of three tones (T1, T2, T3). The frequency of T1 varied from trial to trial and T3 could be higher, lower, or identical to T1. T2 had to be ignored and the task was to identify the relation between T3 and T1 using three response categories: "up," "down," and "same." Confidence ratings were also collected, in order to measure receiver operating characteristics (ROCs; see, e.g., Green and Swets, 1974, chap. 2). The obtained ROCs were not consistent with the idea that listeners based their judgments on observations made on a single subjective dimension, analogous to the physical dimension of frequency. It appeared instead that an additional subjective dimension was the relative "familiarity" of T3, a variable depending on the *unsigned* frequency difference between T1 and T3. On the latter dimension, "same" trials could be discriminated from both "up" or "down" trials, but "up" trials could not be discriminated from "down" trials. Our own experiment was comparable to Wickelgren's experiment when the elements of *C* were asynchronous. In this case, it can therefore be thought that our listeners were able to use, in the present/absent task, a perceptual cue—"familiarity"—which was not helpful in the up/down task.

Experiment 2

As explained in the Introduction, the main goal of Experiment 2 was to clarify the mechanism enabling listeners to be successful in the up/down task when the elements of *C* are synchronous. We wished to test, more specifically, a hypothesis that can be called the "partial forward masking (PFM)" hypothesis. According to the PFM hypothesis, when the up/down task has to be performed on a *C*-*T* or *T*-*C* sequence, the first of these two stimuli is ignored and listeners consider only the second stimulus; this is sufficient because the second stimulus is partially masked by the first one in different ways on "up" and "down" trials; due to this differential masking effect, the second stimulus is perceptually different on "up" and "down" trials, and listeners use the corresponding cue to identify the direction of the frequency shift. Listeners had the opportunity to learn to use appropriately the hypothetical forward masking cue during the preliminary training phase of our experiments, in which trial-by-trial feedback was provided.

A simple way to test the PFM hypothesis is to insert, on both "up" and "down" trials, a relatively loud burst of wide-band noise between the two stimuli. Given that wide-band noise can mask a tone of any frequency and that the strength of forward masking is an increasing function of the temporal proximity of the masker and the target (Zwislocki et al., 1959), the final stimulus in the *C*-Noise-*T* or *T*-Noise-*C* sequence should no longer be masked by the initial stimulus. Thus, the PFM hypothesis predicts that insertion of noise will dramatically impair listeners' performance if up/down judgments are based on a masking-like interaction between *T* and *C*, rather than on FSDs.

In Experiment 2, acccordingly, the up/down task was performed using *C*-*T* sequences that included, on some trials, an interfering noise burst between *C* and *T*. The noise burst was replaced by silence on other trials, and by a single interfering pure tone in a third set of trials. The interfering pure tone used on a given trial always had a relatively low SPL, and it was always remote in frequency from both *T* and the chord element close to *T*. Thus, this interfering tone was not expected to mask *T*, and the PFM hypothesis predicted that its effect on performance would be smaller than the effect of a noise burst. On the other hand, the interfering tone was expected to be processed by the FSDs, whereas this was not the case for the noise bursts (given that the latter stimuli were not periodic). Moreover, since Experiment 1 led us to suppose that the FSDs are primarily sensitive to the frequency relations of *immediately* consecutive tones, the interfering tone was expected to disrupt the FSDs' sensitivity to the relation between *T* and *C*. Therefore, under the hypothesis that success in the up/down task rests on the use of automatic FSDs, it was predictable that the interfering tones would be more deleterious than interfering noise, contrary to the prediction of the PFM hypothesis.

Method

The *C* sounds used in this experiment consisted of seven synchronous pure tones, spaced in frequency by equal intervals of 750 cents. On each trial, *C* was randomly positioned between 125 and 4000 Hz, using a logarithmic frequency scaling. The pure tone *T* presented after *C* was positioned at random 100 cents above or below one of the elements of *C*, selected at random, and the task was to identify the direction of this frequency shift. *T* and each element of *C* had a nominal SPL of 65 dB.

Between *C* and *T*, there was a 1-s ISI containing either an interfering noise burst (*Noise* condition), an interfering pure tone (*Tone* condition), or no stimulus at all (*None* condition). Each stimulus (*C*, *T*, or interfering stimulus) had a total duration of 300 ms and was gated on and off with 20-ms raised-cosine amplitude ramps. When an interfering stimulus was presented, the ISIs separating it from *C* and from *T* had equal durations (350 ms).

In the *Noise* condition, the presented noise bursts consisted of pink noise, generated with the algorithm described by Gardner (1978). A new noise sample was generated on each trial. It was presented at a level of 72 dB (A weighting), and therefore its loudness was approximately the same as the loudness of *C*.

In the *Tone* condition, the frequency of the interfering tone varied randomly from trial to trial, between 125 and 4000 Hz. However, the interfering tone was constrained to be positioned at least 300 cents away from the crucial element of *C*, i.e., the element close to *T*. We wished to present the interfering tones at an intensity level at which they would be just-detectable if they were mixed with the noise bursts of the *Noise* condition. To this end, each of the listeners participating in Experiment 2 initially had to perform a tone-in-noise detection task. On each trial, two successive noise bursts similar to those used in the main part of the experiment were presented. A pure tone with a randomly selected frequency (between 125 and 4000 Hz) was added to either the first or the second noise burst (at random), and the listener had to identify the temporal position of the corresponding noise burst. The pure tone and its masking noise burst were gated on and off synchronously. Feedback concerning response accuracy was provided visually following each trial. From trial to trial, the SPL of the pure tone was varied adaptively, using the "weighted up-down" procedure described by Kaernbach (1991), in order to estimate a detection threshold defined as the SPL for which the probability of a correct response was 0.75. Subsequently, in the *Tone* condition of the experiment, the interfering tones presented to a given listener were at the SPL found to be the detection threshold for this listener. The SPLs in question had a mean value of 56 dB and ranged from 55 to 59 dB.

In all three interference conditions—*Noise*, *Tone*, and *None*—a random melody of five pure tones was presented at the beginning of every trial, exactly as in Experiment 1. Listeners were not provided with immediate feedback after a trial, but were allowed to examine their results following each block of trials. Trials were organized in blocks of 50, during which the interference condition was fixed. In a given experimental session, the three conditions were run once, in a random order. For each listener, twelve such sessions were run, thus providing a total of 600 trials per condition. The equipment used was the same as in Experiment 1, and the stimuli were again presented diotically.

Five listeners with normal hearing (four men, one woman) were tested. This group included four students in their twenties and the first author. All of these listeners were amateur musicians. Two of them had previously participated in related experiments, whereas the other three had no previous experience with psychoacoustics. For the naive listeners, the experiment proper was preceded by two or three training sessions.

Results

Performance was again measured in terms of d'. As regards the *None* or *Noise* condition, we simply computed a single d' statistic from the 600 trials run for each listener. The *Tone* condition was treated differently. In that case, it could be expected that listeners' judgments ("up" or "down") would be biased by the direction of the frequency shift taking place between the interfering tone and *T*. Therefore, separate d' statistics were computed from the trials on which the interfering tone had been, respectively, lower and higher than *T*; the two d' values obtained for each listener were then averaged. However, the outcome of this analysis (a global d' of 0.88) was very similar to that obtained when the data were processed as in the other two conditions (global d': 0.84). Moreover, it appeared that listeners' decision criteria [log(β)] were in fact not influenced in a statistically significant way by the direction of the frequency shift between the interfering tone and *T*, $t(4) = 1.39$, $p > .10$.

The individual and mean d' values obtained in all three conditions are displayed in Figure 3. It can be seen that mean performance was barely poorer in the *Noise* condition than in the *None* condition, but definitely poorer in the *Tone* condition than in the other two conditions. A repeated-measures ANOVA of the individual data showed that the overall effect of interference condition was significant, $F(2, 8) = 13.34$; $p = .003$, $\eta^2 = 0.77$. Tukey-Kramer multiple-comparison tests (with alpha set to 0.05) indicate

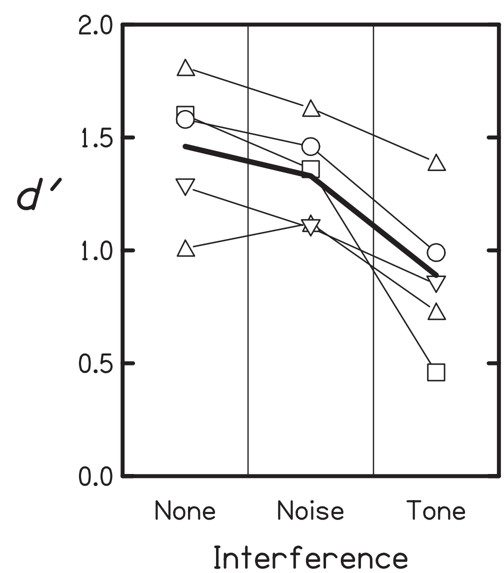

Figure 3. Results obtained in the three conditions (*None*, *Noise*, *Tone*) of Experiment 2. Open symbols represent the individual data. Thick lines connect the mean data.

that d' was significantly different in the *None* and *Tone* conditions, as well as in the *Noise* and *Tone* conditions, but not in the *None* and *Noise* conditions.

Discussion

The experimental results clearly contradict the PFM hypothesis. As explained above, this hypothesis predicted that the *Noise* condition would be more difficult than the *Tone* condition. We found just the opposite. The interfering noise bursts used in the *Noise* condition failed to make performance significantly poorer than in the *None* condition. This is remarkable because these noise bursts were as loud as the *C* stimuli, and were indeed intense enough to mask almost completely the pure tones that produced a strong interference effect in the *Tone* condition.

In contrast, the results are consistent with the idea that performance was determined by implicit frequency comparisons stemming from the existence of automatic FSDs. Under that hypothesis, it was predictable that the interfering stimuli used in the *Noise* condition would not impair performance since, being devoid of any periodicity, they were unlikely to be processed by the FSDs. In addition, one can account for the difficulty of the *Tone* condition by assuming that the FSDs are more sensitive to the frequency relation of immediately consecutive tones than to the frequency relation of two tones separated by a third tone. We had already made the same assumption to account for one aspect of the results of Experiment 1, namely the fact that making the elements of *C* asynchronous rather than synchronous did not improve performance in the up/down task. Since in Experiment 1 as well as in the *Tone* condition of Experiment 2, the interfering tones were always positioned at least 300 cents away from *T*, the FSDs were not expected to respond in a strong manner to the frequency relation of *T* and an interfering tone. However, significant responses to such relations were nonetheless predictable because the "tuning" of the FSDs—that is, their relative sensitivity to frequency shifts of various sizes—seems to be rather broad for tone sequences that are not very rapid (Demany et al., 2009). Moreover, it is likely that the FSDs respond more strongly to tones closer in time, and *T* was closer in time to an interfering tone than to the critical element of *C*.

As an alternative explanation of listeners' difficulties in the *Tone* condition of Experiment 2, one might suppose that the elements of *C* could be heard out explicitly and that the interfering tone was deleterious because it disrupted the memory trace of the critical element of *C*. But the crucial part of this hypothesis, i.e., the assumption that the elements of *C* could be heard out explicitly, is untenable. This is shown by a comparison of the results obtained in the *None* condition of the experiment discussed here to the results of three previous experiments (Demany & Ramos, 2005, Experiment 1; Demany et al., 2009, Experiment 1; Demany et al., 2010, Experiment 1). In these three previous experiments, listeners performed not only an up/down task but also a present/absent task. As in the experiment discussed here, *C* consisted of 300-ms synchronous pure tones, *T* was presented after *C* rather than before it, and the frequency shift to be judged in the up/down task had a magnitude of 100 cents. Moreover, *C* contained on average 7 tones, i.e., exactly the number of tones used here, and these tones were on average spaced by intervals of 667 cents, a figure close to the 750 cents used here. In the up/down task of our three previous experiments, d' had an average value of 2.18, while we obtained here a lower mean d', 1.46. This difference may chiefly originate from the fact that the ISI separating *C* from *T* was longer here than in the previous experiments (see in this respect Demany & Ramos, 2005; Demany, Trost, Serman, & Semal, 2008). But more importantly, the average value of d' in the present/absent task of the previous experiments was only 0.41. This strongly suggests that the elements of *C* could not (or only very rarely) be heard out explicitly in the previous experiments, and therefore here also.

General Discussion

Our study clearly dissociated two types of perceptual comparison between the frequencies of temporally separate tones: on one hand explicit comparisons between consciously audible tones, and on the other hand implicit comparisons that can be made even if one of the compared tones is not consciously audible. Experiment 1 demonstrates that the mechanisms underlying explicit and implicit comparisons have different characteristics. The mechanism of explicit comparisons, which is more effective in the present/absent task than in the up/down task, appears to be primarily sensitive to the magnitude of a frequency change. For this mechanism, large changes are more detectable than smaller changes. In contrast, the mechanism of implicit comparisons, which is more effective in the up/down task than in the present/absent task, appears to be primarily sensitive to the direction of small frequency changes. However, our results do not imply that the explicit comparison mechanism ignores change direction. Indeed, performance for explicit comparisons in the up/down task was well above chance, showing that change direction was accessible to the listeners. Nor do our results imply that the implicit mechanism is totally insensitive to change magnitude: We have shown previously that small changes are better detected than large changes for implicit comparisons (Demany et al., 2009). Our experimental paradigm specifically aimed at dissociating the two mechanisms, by introducing intervening tones that disrupted implicit comparisons when explicit access to the tones was possible. In the general case, it is likely that both implicit and explicit comparisons can be combined to produce an accurate encoding of pitch sequences (Dowling & Fujitani, 1971; Cousineau et al., 2009).

Experiment 2 shows that what we call an "implicit comparison" between two tones cannot be reinterpreted, more parsimoniously, as a PFM interaction—a partial masking of the second tone by the first tone. Admittedly, although the results of Experiment 2 rule out the PFM hypothesis, they do not rule out the possibility of a stimulus interaction basically different from masking. Masking is chiefly a low-level sensory effect determined by the spectral relations of the masker and target rather than by their perceptual qualities per se (see, e.g., Moore, 1995). For example, a 1-kHz tonal target is easier to mask with wideband noise than with a 1.5-kHz tone, although the latter masker is perceptually more similar to the target. In Experiment 2, therefore, a possible masking effect of *C* on the encoding of *T* was more liable to be cancelled by interfering wide-band noise than by an interfering tone remote in frequency from *T*. Nevertheless, the encoding of a tone at a high level of the auditory system might be more strongly affected by a previous tone than by a previous noise, even when the previous tone is not consciously audible and is a less effective masker than the noise. After all, such a high-level forward interaction could justifiably be called an implicit comparison. Alternatively, the implicit comparisons documented here reflect the activity of automatic FSDs that do not affect the encoding of the tones linked by these FSDs.

Whatever their precise mechanism, implicit as well as explicit frequency comparisons between successive tones require some form of sensory memory (Demany & Semal, 2008). They probably recruit separate memory systems. It is reasonable to suppose, in addition, that the memory system on which implicit comparisons are based has an essentially unlimited capacity whereas this is not the case for the system involved in explicit comparisons. Demany et al. (2008) provided evidence for the former point. In their study, a chord made up of a variable number of pure tones (up to 12 tones) was followed by a single tone after a variable ISI (0 – 2000 ms), and listeners had to perform the up/down task used again here. Performance decreased as the number of tones in the chord increased and as the ISI increased, but there was no interaction between these two factors; thus, listeners' ability to memorize the chord after its offset appeared to be independent of the number of its elements. In contrast, Visscher, Kaplan, Kahana, & Sekuler (2007) have recently provided evidence that explicit short-term auditory memory has a very limited capacity. Therefore, the auditory memory systems underlying implicit and explicit frequency comparisons may to some extent differ from each other in the same way as "iconic" visual memory and "short-term visual memory" differ (Phillips, 1974).

The concept of implicit memory is of course not new. It has been previously invoked in the context of both behavioral studies and physiological studies. In the psychological literature about perception, this concept is typically associated with "priming" (Schacter, 1994) or perceptual learning (Agus, Thorpe, & Pressnitzer, 2010) phenomena. In the physiological domain, Miller and Desimone (1994) have associated it with an effect that they called "adaptive mnemonic filtering" and that they observed in the inferior temporal cortex of monkeys presented with a sequence of visual stimuli. Adaptive mnemonic filtering manifests itself as a decrease in a neuron's response to a given stimulus when the same stimulus has already been presented before. Miller and Desimone also found that when monkeys have to detect repetitions of a specific stimulus, some neurons in their inferior temporal cortex do not show reduced responses to the repetitions of this stimulus, but on the contrary enhanced responses. This was interpreted as the manifestation of an explicit and active form of memory, contrasting with the implicit and passive form expressed by adaptive mnemonic filtering. Adaptation phenomena akin to those observed by Miller and Desimone have been recently uncovered in the primary auditory cortex of cats (Ulanovsky, Las, & Nelken, 2003) and even in subcortical auditory stations (Anderson, Christianson, & Linden, 2009; Malmierca, Cristaudo, Perez-Gonzalez, & Covey, 2009).

It seems clear, however, that the implicit form of auditory memory documented in the current study is unrelated to the behavioral effects labeled up to now as "priming," as well as to the neural adaptation phenomena that we just mentioned. Both priming and adaptation are maximal when the source stimulus (i.e., the prime or the adaptor) is identical to the target stimulus; the priming or adaptation effect decreases monotonically when the two stimuli are made more and more different. Thus, if an effect of this type could be exploited when a tone has to be compared with the elements of a chord, it would be helpful in the present/absent task but essentially useless in the up/down task. We found on the contrary that implicit frequency comparisons favor the up/down task rather than the present/absent task.

Zatorre and Samson (1991) reported neuropsychological results that apparently fit in with our distinction between implicit and explicit frequency comparisons. In their study, patients with brain lesions in the temporal or frontal lobe had to make same/different judgments on pairs of tones separated by a 1650-ms ISI. On trials for which the correct response was "different," the second tone was higher in frequency than the first tone. The ISI was either silent or occupied by interfering tones varying in frequency. When the ISI was silent, patients' performance did not differ significantly from that of normal controls. In the presence of interfering tones, however, patients with damage in the right temporal lobe or the right frontal lobe showed a significant deficit. The interfering tones presumably prevented the participants from making implicit frequency comparisons. In the absence of interfering tones, on the other hand, implicit (as well as explicit) frequency comparisons were possible. Moreover, the implicit comparisons were potentially helpful, even though participants had to make same/different judgments. This is indeed suggested by results that we have mentioned earlier in the present paper, those obtained in the present/close task of Demany and Ramos (2005). Overall, therefore, the findings of Zatorre and Samson suggest that implicit and explicit frequency comparisons recruit separate brain areas.

Implicit frequency comparisons are in our view performed by FSDs primarily intended to *bind* automatically successive sounds, as already argued in the Introduction. However, successive sounds can be perceptually bound even when they are identical, and it may indeed seem that this complete absence of change is optimal for perceptual binding. Yet, we have supposed that the FSDs are maximally sensitive to frequency shifts of about one tenth of an octave, which means that an absence of shift does not excite them optimally. At first sight, this seems inconsistent with the idea that the FSDs play a major role in perceptual binding. But in fact each FSD might well respond significantly to a sequence of identical tones. Moreover, such a sequence would then excite significantly, to the same extent, the two putative classes of FSDs, respectively tuned to upward and downward shifts. Therefore, if binding per se (i.e., the subjective coherence of the sequence) is determined by the *sum* of the responses of all FSDs, then it can be substantial for a sequence of two identical tones. Many more studies are needed to clarify the FSDs' precise properties and their role in auditory scene analysis.

References

Agus, T. A., Thorpe, S. J., & Pressnitzer, D. (2010). Rapid formation of auditory memories: Insights from noise. *Neuron, 66,* 610–618.

Anderson, L. A., Christianson, G. B., & Linden, J. F. (2009). Stimulus-specific adaptation occurs in the auditory thalamus. *Journal of Neuroscience, 29,* 7359–7363.

Bernstein, J. G., & Oxenham, A. J. (2003). Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number? *Journal of the Acoustical Society of America, 113,* 3323–3334.

Bernstein, J. G., & Oxenham, A. J. (2006). The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level. *Journal of the Acoustical Society of America, 120,* 3916–3928.

Bernstein, J. G., & Oxenham, A. J. (2008). Harmonic segregation through mistuning can improve fundamental frequency discrimination. *Journal of the Acoustical Society of America, 124,* 1653–1667.

Bregman, A. S. (1990). *Auditory scene analysis.* Cambridge, MA: MIT Press.

Cousineau, M., Demany, L., & Pressnitzer, D. (2009). What makes a melody: The perceptual singularity of pitch sequences. *Journal of the Acoustical Society of America, 126,* 3179–3186.

Darwin, C. J. (1984). Perceiving vowels in the presence of another sound: Constraints on formant perception. *Journal of the Acoustical Society of America, 76,* 1636–1647.

Demany, L., Pressnitzer, D., & Semal, C. (2009). Tuning properties of the auditory frequency-shift detectors. *Journal of the Acoustical Society of America, 126,* 1342–1348.

Demany, L., & Ramos, C. (2005). On the binding of successive sounds: Perceiving shifts in nonperceived pitches. *Journal of the Acoustical Society of America, 117,* 833–841.

Demany, L., & Semal, C. (2008). The role of memory in auditory perception. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 77–113). New York: Springer.

Demany, L., Semal, C., Cazalets, J. R., & Pressnitzer, D. (2010). Fundamental differences in change detection between vision and audition. *Experimental Brain Research, 203,* 261–270.

Demany, L., Trost, W., Serman, M., & Semal, C. (2008). Auditory change detection: Simple sounds are not better memorized than complex sounds. *Psychological Science, 19,* 85–91.

Deutsch, D. (1972). Mapping of interactions in the pitch memory store. *Science, 175,* 1020–1022.

Deutsch, D. (1999). The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music* (pp. 349–411). New York: Academic Press.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49,* 524–531.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, S., & Kidd, G., Jr. (2003). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *Journal of the Acoustical Society of America, 114,* 368–379.

Elliott, L. L. (1971). Backward and forward masking. *Audiology, 10,* 65–76.

Gardner, M. (1978). White and brown music, fractal curves and 1/f fluctuations. *Scientific American, 238,* 16–31.

Gentner, T. Q. (2008). Temporal scales of auditory objects underlying birdsong vocal recognition. *Journal of the Acoustical Society of America, 124,* 1350–1359.

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics.* Huntington, New York: Krieger.

Helmholtz, H. L. F. von (1859). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik.* Braunschweig, Germany: Vieweg & Sohn. [English translation of the fourth German edition (1877) *On the sensations of tone as a physiological basis for the theory of music,* by A. J. Ellis, 1954, New York: Dover.]

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception and Psychophysics, 49,* 227–229.

Kidd, G. D., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational masking. In: W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 143–189). New York: Springer.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge, MA: Cambridge University Press.

Malmierca, M. S., Cristaudo, S., Perez-Gonzalez, D., & Covey, E. (2009). Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. *Journal of Neuroscience, 29,* 5483–5493.

Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science, 263,* 520–522.

Moore, B. C. J. (1995). Frequency analysis and masking. In B. C. J. Moore (Ed.), *Hearing* (pp. 161–205). San Diego: Academic Press.

Moore, B. C. J., Glasberg, B. R., & Jepsen, M. L. (2009). Effects of pulsing of the target tone on the audibility of partials in inharmonic complex tones. *Journal of the Acoustical Society of America, 125,* 3194–3204.

Moore, B. C. J., & Ohgushi, K. (1993). Audibility of partials in inharmonic complex tones. *Journal of the Acoustical Society of America, 93,* 452–461.

Neff, D. L. (1991). Forward masking by maskers of uncertain frequency content. *Journal of the Acoustical Society of America, 89,* 1314–1323.

Neff, D. L. (1995). Signal properties that reduce masking by simultaneous, random-frequency maskers. *Journal of the Acoustical Society of America, 98,* 1909–1920.

Neff, D. L., & Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Perception and Psychophysics, 41,* 409–415.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics, 16,* 283–290.

Roberts, B., & Bregman, A. S. (1991). Effects of the pattern of spectral spacing on the perceptual fusion of harmonics. *Journal of the Acoustical Society of America, 90,* 3050–3060.

Schacter, D. L. (1994). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. In D. L. Schacter & E. Tulving (Eds.), *Memory systems,* 1994 (pp. 233–268). Cambridge, MA: MIT Press.

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience, 6,* 391–398.

Visscher, K. M., Kaplan, E., Kahana, M. J., & Sekuler, R. (2007). Auditory short-term memory behaves like visual short-term memory. *PloS Biology, 5,* e56.

Wickelgren, W. A. (1969). Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology, 6,* 13–61.

Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M., & Neiworth, J. J. (2000). Music perception and octave generalization in rhesus monkeys. *Journal of Experimental Psychology: General, 129,* 291–307.

Zatorre, R. J., & Samson, S. (1991). Role of the right temporal neocortex in retention of pitch in auditory short-term memory. *Brain, 114,* 2403–2417.

Zwislocki, J., Pirodda, E., & Rubin, H. (1959). On some poststimulatory effects at the threshold of audibility. *Journal of the Acoustical Society of America, 31,* 9–14.

Received July 9, 2009
Revision received March 9, 2010
Accepted March 29, 2010 ■